

PCT

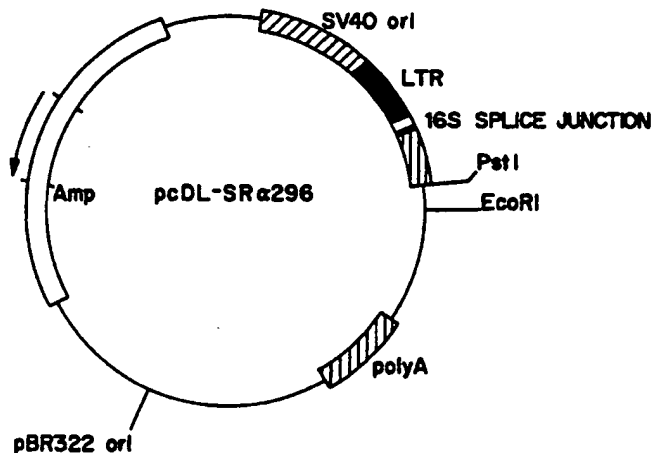
WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

| | | | |
|---|--|--|---|
| (51) International Patent Classification 6 : C12Q 1/02, 1/37, C12N 5/10, 15/25 | | A1 | (11) International Publication Number: WO 96/34976 |
| | | | (43) International Publication Date: 7 November 1996 (07.11.96) |
| (21) International Application Number: PCT/US96/06070 | | (81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). | |
| (22) International Filing Date: 1 May 1996 (01.05.96) | | | |
| (30) Priority Data: 08/432,693 1 May 1995 (01.05.95) US | | | |
| (71) Applicant: VERTEX PHARMACEUTICALS INCORPORATED [US/US]; 40 Allston Street, Cambridge, MA 02139-4211 (US). | | | |
| (72) Inventor: SU, Michael; 15 Donna Road, Newton, MA 02159 (US). | | Published With international search report. | |
| (74) Agents: HALEY, James, F., Jr. et al.; Fish & Neave, 1251 Avenue of the Americas, New York, NY 10020-1104 (US). | | | |

(54) Title: METHODS, NUCLEOTIDE SEQUENCES AND HOST CELLS FOR ASSAYING EXOGENOUS AND ENDOGENOUS PROTEASE ACTIVITY



(57) Abstract

The invention relates to methods for assaying exogenous protease activity in a host cell transformed with nucleotide sequences encoding that protease and a specialized substrate. It also relates to methods for assaying endogenous protease activity in a host cell transformed with nucleotide sequences encoding a specialized substrate. When these nucleotide sequences are expressed, the exogenous or endogenous protease cleaves the substrate and releases a polypeptide that is secreted out of the cell, where it can be easily quantitated using standard assays. The methods and transformed host cells of this invention are particularly useful for identifying inhibitors of the exogenous and endogenous proteases. If the protease is a protease from an infectious agent, inhibitors identified by these methods are potential pharmaceutical agents for the treatment or prevention of infection by that agent.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | |
|----|--------------------------|----|--|----|--------------------------|
| AM | Armenia | GB | United Kingdom | MW | Malawi |
| AT | Austria | GE | Georgia | MX | Mexico |
| AU | Australia | GN | Guinea | NE | Niger |
| BB | Barbados | GR | Greece | NL | Netherlands |
| BE | Belgium | HU | Hungary | NO | Norway |
| BF | Burkina Faso | IE | Ireland | NZ | New Zealand |
| BG | Bulgaria | IT | Italy | PL | Poland |
| BJ | Benin | JP | Japan | PT | Portugal |
| BR | Brazil | KE | Kenya | RO | Romania |
| BY | Belarus | KG | Kyrgyzstan | RU | Russian Federation |
| CA | Canada | KP | Democratic People's Republic of Korea | SD | Sudan |
| CF | Central African Republic | KR | Republic of Korea | SE | Sweden |
| CG | Congo | KZ | Kazakhstan | SG | Singapore |
| CH | Switzerland | LI | Liechtenstein | SI | Slovenia |
| CI | Côte d'Ivoire | LK | Sri Lanka | SK | Slovakia |
| CM | Cameroon | LR | Liberia | SN | Senegal |
| CN | China | LT | Lithuania | SZ | Swaziland |
| CS | Czechoslovakia | LU | Luxembourg | TD | Chad |
| CZ | Czech Republic | LV | Latvia | TG | Togo |
| DE | Germany | MC | Monaco | TJ | Tajikistan |
| DK | Denmark | MD | Republic of Moldova | TT | Trinidad and Tobago |
| EE | Estonia | MG | Madagascar | UA | Ukraine |
| ES | Spain | ML | Mali | UG | Uganda |
| FI | Finland | MN | Mongolia | US | United States of America |
| FR | France | MR | Mauritania | UZ | Uzbekistan |
| GA | Gabon | | | VN | Viet Nam |

METHODS, NUCLEOTIDE SEQUENCES AND HOST CELLS FOR
ASSAYING EXOGENOUS AND ENDOGENOUS PROTEASE ACTIVITY

TECHNICAL FIELD OF INVENTION

5 The invention relates to methods for assaying
exogenous protease activity in a host cell transformed
with nucleotide sequences encoding that protease and a
specialized substrate. It also relates to methods for
assaying endogenous protease activity in a host cell
10 transformed with nucleotide sequences encoding a
specialized substrate. When these nucleotide sequences
are expressed, the exogenous or endogenous protease
cleaves the substrate and releases a polypeptide that is
secreted out of the cell, where it can be easily
15 quantitated using standard assays. The methods and
transformed host cells of this invention are particularly
useful for identifying inhibitors of the exogenous and
endogenous proteases. If the protease is a protease from
an infectious agent or is characteristic of a diseased
20 state, inhibitors identified by these methods are
potential pharmaceutical agents for treatment or
prevention of the disease.

BACKGROUND ART

Proteases play an important role in the
25 regulation of many biological processes. They also play
a major role in disease. In particular, proteolysis of
primary polypeptide precursors is essential to the
replication of several infectious viruses, including HIV
and HCV. These viruses encode proteins that are
30 initially synthesized as large polyprotein precursors.
Those precursors are ultimately processed by the viral
protease to mature viral proteins. In light of this,
researchers have begun to concentrate on inhibition of
viral proteases as a potential treatment for certain
35 viral diseases.

Proteases also play a role in non-infectious

-2-

diseases. For example, changes in normal cellular function may cause an undesirable increase or decrease in proteolytic activity. This often leads to a disease state.

5 The ability to detect viral or mutant protease activity in a quick and simple assay is important in the biochemical characterization of these proteases and in the screening and identification of potential inhibitors. Several of these assays have been described in the art.

10 T. M. Block et al., Antimicrob. Agents Chemother., 34, pp. 2337-41 (1990) described a prototype assay for screening potential HIV protease inhibitors. This assay involved cloning the HIV protease recognition sequence into the tetracycline resistance gene (Tet^R) of pBR322 and cotransforming *E. coli* with the modified Tet^R gene and the gene encoding the HIV protease. Co-expression of these two genes caused tetracycline sensitivity. Potential inhibitors were identified by the ability to restore tetracycline resistance to the
20 transformed bacteria.

 E. Sarubbi et al., FEBS Lett., 279, pp. 265-69 (1991) described another assay for detecting HIV protease inhibitors that utilized a HIV-1 Gag- β -galactosidase fusion protein and a monoclonal antibody that bound to
25 the fusion protein in the gag region. Coexpression of the HIV protease and the fusion protein lead to cleavage of the latter and abolished monoclonal antibody binding. Potential inhibitors were identified by increased binding of the monoclonal antibody to the fusion protein.

30 T. A. Smith et al., Proc. Natl. Acad. Sci. USA, 88, pp. 5159-62 (1991), B. Dasmahapatra et al., Proc. Natl. Acad. Sci. USA, 89, pp. 4159-62 (1992) and M. G. Murray et al., Gene, 134, pp. 123-28 (1993) each described protease assay systems utilizing the yeast GAL4
35 protein. Each of these authors described inserting a protease cleavage site in between the DNA binding domain and the transcriptional activating domain of GAL4.

-3-

Cleavage of that site by a coexpressed protease renders GAL4 transcriptionally inactive leading to the inability of the transformed yeast to metabolize galactose.

- H.-D. Liebig et al., Proc. Natl. Acad. Sci. USA, 88, pp. 5979-83 (1991) disclosed the use of a fusion protein consisting of a self-cleaving protease fused to the α fragment of β -galactosidase to assay protease activity. Active forms of the protease cleaved themselves off of the fusion protein and the resulting protein was able to carry out α -complementation. Fusions containing inactive protease were unable to perform α -complementation.

- Y. Komoda et al., J. Virol., 68, pp. 7351-57 (1994) described an assay to identify HCV protease cleavage sites within the HCV precursor polyprotein. These authors created chimeric proteins comprising various portions of the HCV precursor polyprotein inserted in between the *E. coli* maltose binding protein and dihydrofolate reductase. If the HCV portion of these chimeras contained a cleavage site, the chimera would be cleaved when it was coexpressed with HCV protease in *E. coli*. Cleavage of the chimera was determined by SDS-polyacrylamide gel electrophoresis of *E. coli* lysates.

- Y. Hirowatari et al., Anal. Biochem., 225, pp. 113-120 (1995) described another assay to detect HCV protease activity. In this assay, the substrate, HCV protease and a reporter gene are cotransfected into COS cells. The substrate is a fusion protein consisting of (HCV NS2)-(DHFR)-(HCV NS3 cleavage site)-Tax1. The reporter gene is chloramphenicol transferase (CAT) under control of the HTLV-1 long terminal repeat (LTR) and resides in the cell nucleus following expression. The uncleaved substrate is expressed as a membrane-bound protein on the surface of the endoplasmic reticulum due to the HCV NS2 portion. Upon cleavage, the released Tax1 protein translocates to the nucleus and activates CAT expression by binding to the HTLV-1 LTR. Protease

-4-

activity is determined by measuring CAT activity in a cell lysate.

Despite these developments, no one has yet developed a protease assay system that can be carried out with higher eukaryotic cells and is both quantitative and does not require cell lysis prior to quantitation. Avoiding cell lysis prior to quantitation is desirable in that the assay may be performed more rapidly and with less manipulation. Also, lysis can often lead to aberrant results. Thus, there is a need for an accurate and quantitative cellular-based protease assay that can be carried out in a higher eukaryotic cell without cell lysis.

SUMMARY OF THE INVENTION

The present invention fulfills this need by providing methods for assaying exogenous protease activity in a host cell expressing that protease. The methods involve utilizing a host cell expressing a first nucleotide sequence encoding an exogenous protease and a second nucleotide sequence encoding an artificial substrate for that protease. The artificial substrate comprises a cleavage site for the protease situated at or near the natural maturation site of a pre-polypeptide, part of which is secreted following proteolytic processing. When the host is grown under conditions that cause expression of the first and second nucleotide sequences, the exogenous protease cuts the artificial substrate at the cleavage site, releasing the mature polypeptide which is secreted into the growth media. The growth media is then isolated and assayed for the mature polypeptide.

Alternatively, the invention may be utilized to assay endogenous proteases, especially when quantitation of those proteases is difficult due to the inability to detect or distinguish between the cleaved and uncleaved native substrate.

SUBSTITUTE SHEET (RULE 26)

-5-

According to one aspect of the invention, the assay is used to quantitate an exogenous viral protease. Such assays are particularly useful as replacements for current viral protease assays that require the use of intact, infectious virus or where no simple viral model is available to detect viral protease activity. These assays may be used to identify and assay potential inhibitors of viral proteases which, in turn, may be used as pharmaceutical agents for the treatment or prevention of viral disease.

This invention also provides host cells transformed with nucleotide sequences encoding an endogenous protease and a corresponding substrate, as well as those transformed with a specialized substrate for an endogenous protease. These hosts may be used in the methods of this invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 depicts the structure of pcDL-SR α 296.

Figure 2 depicts the structure of a derivative of pKV containing the pre-IL-1 β coding sequence.

Figure 3, panel A, is an immunoblot of cell lysates from cells transfected with a NS3-wild-type or NS3-mutant NS3-4A-4B-IL1 β or cotransfected with a NS3-mutant NS3-4A-4B-IL1 β and a NS3(1-180) construct probed with an anti-NS3 antibody. Figure 3, panel B, is an immunoblot of the same cell lysates probed with an anti-IL-1 β antibody.

Figure 4 depicts the immunoprecipitation of the media from ³⁵S-labelled cells transfected with either a NS3-wild-type or NS3-mutant NS3-4A-4B-IL1 β construct with an anti-IL-1 β antibody.

Figure 5 is an immunoblot of cell lysates from cells co-transfected with NS3-4A and either a NS5A/5B- or CSM-containing pre-IL1 β substrate probed with an anti-IL-1 β antibody.

Figure 6 depicts the immunoprecipitation of the

-6-

media from ³⁵S-labelled cells co-transfected with NS3-4A and either a NS5A/5B- or CSM-containing pre-IL1 β substrate with an anti-IL-1 β antibody.

Figure 7 depicts the inhibition of HCV NS3 protease cleavage of pre-IL-1 β * by varying concentrations of VH16075 and VH15924.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides a method for assaying exogenous protease activity in a host cell comprising the steps of:

(a) incubating a host cell transformed with a first nucleotide sequence encoding an exogenous protease and a second nucleotide sequence encoding an artificial polypeptide substrate under conditions which cause said exogenous protease and said artificial substrate to be expressed;

wherein said substrate comprises:

(i) a cleavage site for said exogenous protease; and
(ii) a polypeptide that is secreted out of said cell following cleavage by said exogenous protease;

(b) separating said host cell from its growth media under non-lytic conditions; and

(c) assaying said growth media for the presence of said secreted polypeptide.

As used herein, the term "exogenous protease" means a protease not normally expressed by the host cell used in the assay. That term includes full-length proteases that are identical to those found in nature, as well as catalytically active fragments thereof.

The choice of exogenous protease to be assayed is solely dependent upon the decision of the user. The only requirements are that: (1) the specificity of the enzyme in terms of what amino acid residues or sequences it cleaves at be known; (2) the primary structure of at

-7-

least the catalytically active portion of the enzyme be known; and (3) a nucleotide sequence encoding at least an enzymatically active portion of the protease exists or can be made and can be expressed in a heterologous host cell.

According to a preferred embodiment, the exogenous protease is a protease encoded by a pathogenic agent. More preferred is a protease encoded by a pathogenic virus. Most preferably, the exogenous protease is the NS3 protease of hepatitis C virus ("HCV").

HCV NS3 protease is a 70 kilodalton protein that is involved in the maturation of viral polypeptides following infection. It is a serine protease which has a Cys-X or Thr-X substrate specificity. It has also been shown that the protease activity of NS3 resides exclusively in the N-terminal 180 amino acids of the enzyme. Therefore, nucleotide sequences encoding anywhere from the first 180 amino acids of NS3 up to the full length enzyme may be utilized in the methods of this invention. Active fragments of other known proteases may also be used as an alternative to the full-length protease.

According to an alternative embodiment, the invention provides a method for assaying endogenous protease activity in a host cell comprising the steps of:

a) incubating a host cell transformed with a nucleotide sequence encoding an artificial polypeptide substrate under conditions which cause said artificial substrate to be expressed;

wherein said substrate comprises:

i) a cleavage site for said endogenous protease; and
ii) a polypeptide that is secreted out of said cell following cleavage by said endogenous protease;

b) separating said host cell from its growth

-8-

media under non-lytic conditions; and

c) assaying said growth media for the presence of said secreted polypeptide.

5 The term "endogenous protease", as used throughout this application, refers to a proteases that is normally expressed by the host cell. It includes both wild type proteases, as well as naturally occurring mutant proteases with increased or decreased activity.

10 According to the invention, the artificial polypeptide substrate used in the methods must comprise a cleavage site for the protease to be assayed; and must be secreted out of the cell following cleavage by that protease. Preferably, the DNA encoding the artificial substrate is derived from a gene or cDNA encoding a
15 naturally occurring polypeptide that is normally cleaved and then secreted out of a cell, but not necessarily cleaved by the cell utilized in the assay.

The DNA encoding that polypeptide is then modified by inserting, in frame with the polypeptide
20 coding sequence, nucleotides encoding a cleavage site that is recognized by the exogenous protease to be tested. If the cell utilized in the assay is capable of cleaving the substrate at its native cleavage site, then the nucleotides encoding the polypeptide's native
25 cleavage site must be altered so as to render it uncleavable by endogenous proteases.

The protease cleavage site in the artificial substrate is preferably inserted within 60 amino acids on either side of the native cleavage site. Preferably, the
30 artificial cleavage site is inserted N-terminal to the native cleavage site. Alternatively, the protease cleavage site can be created by mutating the native polypeptide sequence. Such mutation is preferably performed on a sequence within 60 amino acids, more
35 preferably N-terminal to the native cleavage site and within 8-10 amino acids of the native cleavage site; or is a mutation of the native cleavage site itself.

-9-

Alteration of the native cleavage site to render it uncleavable by the host cell may be achieved, if necessary, by insertion, deletion or mutation of nucleotides at that site.

5 Insertion of the protease cleavage site into the substrate and alteration of its native cleavage site may be accomplished by any combination of a number of recombinant DNA techniques well known in the art, such as
10 digest/ligation cloning techniques. Alternatively, the DNA encoding all or part of the artificial substrate may be produced synthetically using a commercially available automated oligonucleotide synthesizer. Regardless of the
15 techniques used to insert the protease cleavage site into the substrate polypeptide or alter its native cleavage site, it is crucial that the reading frame of the substrate polypeptide remain intact, without the insertion of stop codons.

 The choice of secretable polypeptide from which
20 the artificial substrate is derived may be selected from any pre-polypeptide that can be cleaved by and the resulting mature polypeptide secreted out of the host cell used for the assay, but is not normally present in that cell. For use in eukaryotic cells there are two
25 main categories of pre-polypeptide from which the choice can be made.

 The first and preferred category comprises pre-polypeptides that are expressed and cleaved in the cytoplasmic compartment. Among these proteins are
30 interleukin-1 β (IL-1 β), interleukin-1 α (IL-1 α), basic fibroblast growth factor (bFGF) and endothelial-monocyte activating polypeptide II (EMAP-II). The advantage of using cytoplasmic pre-polypeptides is that there is a much greater likelihood that the protease and the
35 artificial substrate will share the same subcellular compartment. This is because most proteases of interest are also cytoplasmic proteins and thus will have access

SUBSTITUTE SHEET (RULE 23)

-10-

to the artificial substrate.

The second category of pre-polypeptides that may be used to create artificial substrates used in the methods of this invention are those that are expressed on the cell surface through the organellar secretory pathway and are retained on the cell surface. Such substrates are useful to assay endogenous and exogenous cell membrane proteases, as well as exogenous proteases that are similarly engineered to be cell membrane proteins. The technique of creating a cell membrane protease or substrate involves cloning a leader peptide (i.e., signal sequence) onto the N-terminus of the substrate or protease and a hydrophobic, membrane anchor sequence (either a transmembrane domain or a glycosylphosphatidyl-inositol anchor sequence) onto the C-terminus. The resulting substrate is a cell membrane protein with an extracellularly located cleavage site. When cleaved by a cell membrane protease on the same or a neighboring cell, the secreted polypeptide portion of the substrate is released into the media.

Examples of sequences that may be used for anchoring these proteins in the membrane are the transmembrane domains of TNF α precursor [Nedopsasov et al., Cold Spring Harb. Symp. Quant. Biol., 51, pp. 611-24 (1986)], SP-C precursor [Keller et al., Biochem J., 277, pp. 493-99 (1991)], or alkaline phosphatase [Berger et al., Proc. Natl. Acad. Sci. USA, 86, pp. 1457-60 (1989)].

Techniques for cloning a signal sequence onto a cytoplasmic protein have been well documented [see, for example, Kizer and Trosha, BBRC, 174, pp. 586-92 (1991); Jost et al., J. Biol. Chem., 269, pp. 26267-72 (1994) (expression and secretion of functional single chain Fv molecules using immunoglobulin light chain leader sequence); and Sasada et al., Cell Structure Function, 13, pp. 129-41 (1988) (secretion of human EGF and IgE in mammalian cells using an IL-2 leader sequence)], as have techniques for cloning a transmembrane anchor sequences

-11-

onto cytoplasmic proteins [Berger et al., supra; Oda et al., Biochem J., 301, pp. 577-83 (1984)]. By combining these two techniques, the protease or substrate of interest can be converted from a cytoplasmic protein into
5 a cell surface membrane protein.

In order to insure that the substrate and protease will have access to one another and according to an alternate embodiment of the invention, the artificial substrate and an exogenous protease to be assayed may be
10 encoded as part of a single polyprotein. That polyprotein may be a cytoplasmic or a membrane protein, as long as the substrate and protease domains reside in the same cellular compartment.

The choice of host cell to use in this method is virtually unlimited. Any cell that can grow in
15 culture, be transformed or transfected with heterologous nucleotide sequences and can express those sequence may be employed in this method. These include bacteria, such as E. coli, Bacillus, yeast and other fungi, plant cells,
20 insect cells, mammalian cells. In addition, expression of either of those sequences in higher eukaryotic host cells may be transient or stable. Preferably, the host cell is a higher eukaryotic cell that is incapable of cleaving the substrate at its native cleavage site.
25 Preferably, the host cell is a mammalian cell. Most preferably, the host cell is a COS cell.

It will be apparent that the specific choice of cell is governed by the particular protease to be assayed and by the particular artificial substrate used. In
30 embodiments that assay an exogenous protease, one obvious limitation is that the endogenous cellular enzymes of the chosen host must be unable to cleave the artificial substrate to any significant extent. The endogenous rate of artificial substrate cleavage may be determined by
35 transforming the selected host cell with only the nucleotide sequence coding for the artificial substrate and then growing that host under conditions which cause

-12-

expression of that nucleotide sequence and which would cause expression of the exogenous protease-encoding nucleotide sequence if that sequence were present. The growth media of the cell is then assayed for the presence of the secreted polypeptide portion of the substrate. In assays that measure exogenous protease activity, control cells (no exogenous protease expressed) should secrete less than 10% of the total amount of expressed substrate (due to endogenous cleavage and, in assays that do not distinguish between cleaved and uncleaved substrates, leeching of uncleaved substrate out of the cell) in order to be useful in the methods of this invention. When an endogenous protease is assayed, a controls for non-specific substrate cleavage is a cell transformed with a substrate that contain a mutation at the cleavage site. This mutation renders the substrate uncleavable by the specific endogenous protease being assayed, but still susceptible to non-specific cleavage. As with assays for exogenous proteases, control cells should secrete less than 10% of the total amount of expressed substrate.

In order to quantitate the protease activity, the amount of secreted substrate polypeptide is measured. Quantitation may be achieved by subjecting the growth media to any of the various standard assay procedures that are well known in the art. These include, but are not limited to, immunoblotting, ELISA, immunoprecipitation, RIA, other colorimetric assays, enzymatic assay or bioassay. Quantitation techniques that employ antibodies, preferably utilize antibodies that have low cross-reactivity with the uncleaved substrate. Preferably cross-reactivity is less than 20% and more preferably less than 5%.

According to another embodiment, the present invention provides a method of screening for protease inhibitors. In this method, the above-described assay is carried out in the presence and absence of potential inhibitors of the protease. When the assays of this

SUBSTITUTE SHEET (RULE 26)

-13-

invention are performed using cells which transiently express the substrate and protease, the inhibitor is preferably added immediately after transfection with the protease and substrate-encoding DNA sequences. When
5 stable transformants are used, the potential inhibitor is added at the beginning of the assay. The efficacy of the potential inhibitor (and its ability to cross the cell membrane) is determined by comparing the amount of secreted substrate polypeptide present in the media of
10 cells assayed in its presence versus its absence. Compounds which cause at least a 90% reduction in the amount of secreted substrate polypeptide are potentially useful protease inhibitors.

In order that the invention described herein
15 may be more fully understood, the following examples are set forth. It should be understood that these examples are for illustrative purposes only and are not to be construed as limiting this invention in any manner.

EXAMPLE 1

20 Construction Of Expression Plasmids

A. HCV NS3 Protease

We cloned the nucleotide sequence coding for the entire, intact HCV NS3 protease, an NS3-4A
polyprotein or a truncated NS3 consisting of amino acids
25 1 to 180 into the mammalian expression plasmid pcDL-SRα [Y. Takebe et al., Mol. Cell. Biol., 8, pp. 466-72 (1988)]. That plasmid contains an SV40 origin of replication and an HTLV LTR enhancer/promoter sequence which ultimately drives the high level expression of the
30 NS3 coding sequences (Figure 1).

The respective NS-3 coding fragments (full length NS3, NS3-4A polyprotein or truncated NS3 (amino acids 1-181) were obtained by PCR of the corresponding portions of a full length HCV H strain cDNA (SEQ ID
35 NO:1). For each of the three coding fragments the following 5' primer was used (SEQ ID NO:2):
5'GGACTAGTCTGCAGTCTAGAGCTCCATGGCGCCCATCACGGCGTACG3'. The

-14-

fragment-specific 3' primers used were:

NS3 - (SEQ ID NO:3):

3'GAAGATCTGAATTCTAGATTTTACGTGACGACCTCCACGTCGGC5';

NS3-4A - (SEQ ID NO:4):

5 3'GAAGATCTGAATTCTAGATTTTAGCACTCTTCCATCTCATCGAA5'; and

NS3(1-181) - (SEQ ID NO:5):

3'GAAGATCTGAATTCTAGATTTTAGGATCTCATGGTTGTCTCTAGG5'. These primers produced PCR-amplified fragments containing multiple restriction sites at either end for ease of cloning.

10

In order to ligate the fragments to the vector, the vector was first cleaved with PstI and EcoRI to remove a small fragment. The cut vector was then purified and ligated to the respective PstI/EcoRI cut NS3 protease-encoding fragment.

15

B. IL-1 β /NS3 Substrate

A derivative of plasmid pKV containing the pre-IL-1 β coding sequence has been described by P. K. Wilson et al., *Nature*, 370, pp. 253-70 (1994). That plasmid contains the SV40 origin of replication and the early promoter. The pre-IL-1 β sequence was cloned between the SpeI and BglII sites shown in Figure 2.

20

We inserted a double stranded synthetic DNA fragment (SEQ ID NO:6) which encoded 20 amino acids: SEQ ID NO:7: GADTEDVVCCSMSYTWGVH and contained linkers at both ends that included an ApaI restriction site. The DNA was cloned into the ApaI site in pre-IL-1 β (between the codons for amino acids His₁₁₅ and Asp₁₁₆), immediately upstream of the native cleavage site (located between Asp₁₁₆ and Ala₁₁₇). The first 18 amino acids of the insert correspond to the HCV peptide 5A/5B cleavage site. The last two amino acids are encoded by the linker. The inserted DNA maintained the reading frame of the native pre-IL-1 β protein. The resulting substrate is referred to throughout the application as "pre-IL-1 β *".

30

35

NS3 cleaves the inserted peptide in between the cysteine and serine residues. Because the COS cells we

SUBSTITUTE SHEET (RULE 26)

-15-

utilized in this assay were incapable of cleaving pre-IL-1 β (data not shown), we did not have to knock out the native pre-IL-1 β cleavage site.

In another construct, we performed site directed mutagenesis to alter the native pre-IL-1 β cleavage site of Asp₁₁₆-Ala₁₁₇-Pro₁₁₈ to Cys-Ser-Met, a conserved recognition sequence for NS3. This construct is referred to throughout the application as "pre-IL-1 β B(CSM)".

10 C. NS3-4A-4B-IL-1 β

In order to create a single fusion polypeptide that encoded both the exogenous protease and the polypeptide substrate, we utilized the fact that NS3 can autoprocess (cleave) an NS3-4A-4B polyprotein at both the NS3-4a and 4A-4B junctions.

We isolated a DNA fragment that encoded NS3-4A and the first 60 amino acids of 4B through PCR using the HCV strain H cDNA referred to above (SEQ ID NO:1) and the following primers: SEQ ID NO:8:
20 5'GGACTAGTCTGCAGTCTAGAGCTCCATGGCGCCCATCACGGCGTACG3' and
SEQ ID NO:9: 3'GGACGCGGTCTGCAGGAGGCCGAGGGC5'. The PCR products were digested with PstI and XbaI prior to cloning.

The mature IL-1 β portion of the construct
25 (amino acids 117-269 of SEQ ID NO:11) was created by PCR cloning of full length pre-IL-1 β cDNA (SEQ ID NO:10) using the following primers:

SEQ ID NO:12: 5'CTCGGCCTCCTGCAGGCACCTGTACGATCACTGAAC3';
and SEQ ID NO:13: 3'GGGAATTCTAGATTTTAGGAAGACACAAATTG5'.
30 These PCR products were digested with PstI and EcoRI prior to cloning.

The NS3-4A-4B and IL-1 β fragments were then ligated together with XbaI/EcoRI digested pcDL-SR α to obtain the desired construct.

35 As a control we created a mutant NS3 protease fusion protein construct. This construct was identical to the one described above, except that the NS3 portion

-16-

was created by PCR using the same primers and the cDNA of the NS3 active site mutant S1165A [A. Grakoui et al., J. Virol., 67, pp. 2832-43 (1993)]. The NS3 active site mutant contains a serine-to-alanine mutation in its active site, rendering the enzyme inactive.

EXAMPLE 2

Transfection Of COS Cells And Assay Of Secreted IL-1 β

The expression plasmid constructs described in Example 1 were transfected into COS-7 cells using the DEAE-Dextran transfection protocol [Gu et al., Neuron, 5, pp. 147-57 (1990)]. COS cells in 6-well clusters or 100 mm dishes at 50% confluency were transfected with 4-10 μ g of the desired plasmid in a DEAE-Dextran solution. Following transfection, the cells were incubated an additional 48 hours before assaying.

The processing of pre-IL-1 β or NS3-4A- Δ 4B-IL-1 β fusion protein and subsequent secretion of mature IL-1 β into the media was measured by ELISA of IL-1 β using an antibody that was specific for mature IL-1 β (approx. 3% cross-reactivity with pre-IL-1 β). We analyzed expression by harvesting the COS cells in ice-cold phosphate buffered saline, lysing the cells in a 0.1% Triton X-100 buffer and centrifuging the lysate to remove cell debris. The lysates were then analyzed by SDS-PAGE and immunoblotting using an IL-1 β antibody (Genzyme) and an NS3 antibody. Alternatively, expression, processing and secretion was analyzed by labelling the cells for 24 hours in the presence of [³⁵S]-methionine, incubating the cells for an additional 24 hours after the label was removed and then utilizing immunoprecipitation and SDS-PAGE to analyze the polypeptides.

EXAMPLE 3

NS3-Specific Processing Of An NS3-4A- Δ 4B-IL-1 β Fusion Protein And Secretion Of Δ 4B-IL-1 β Into The Media

Transfectants expressing the NS3-4A- Δ 4B-IL-1 β

-17-

fusion protein autoprocessed that protein at both the NS3-4A and 4A-4B junctions. The cell lysates of these transfectants were subjected to Western blotting utilizing an anti-NS3 antibody. Figure 3, panel A, Wt-1 and Wt-2 lanes, shows that this experiment produced a doublet band in the 70 kD area, present only as a single band in the untransformed control cells (panel A, No DNA lane). The second band of the doublet in the Wt-1 and Wt-2 lanes corresponds to the size of mature NS3. A transfectant that expressed an inactive mutant NS3-containing NS3-4A- Δ 4B-IL-1 β fusion protein demonstrated no 70 kDa doublet and therefore was not autoprocessed (NS3 mutant lane). A transfectant that co-expressed the same mutant fusion protein together with a truncated, but active NS3 -- NS3(1-180) -- was also analyzed. Surprisingly, the mutant fusion protein did not appear to be cleaved by NS3(1-180), as indicated by the lack of a doublet in the 70 kDa region (NS3 mutant + NS3(1-180) lane). However, a 20 kDa band representing the truncated NS3 was detected in that lysate, as indicated by the NS3(1-180) arrow.

A similar experiment performed on cell lysates utilizing an mature IL-1 β -specific antibody demonstrated the presence of a band corresponding in size to the Δ 4B-IL-1 β portion of the fusion protein in both the NS3-4A- Δ 4B-IL-1 β transfectants (Figure 3, panel B, Wt-1 and Wt-2 lanes) and, to a lesser degree in the NS3 mutant fusion protein/NS3(1-180) cotransfectant. Virtually no IL-1 β was detected in the NS3 mutant fusion protein expressing transfectant (IL-1 β arrow). These experiments confirm that the cleavage observed in the wild type NS3-4A- Δ 4B-IL-1 β transfectants was dependent upon NS3 protease activity. Thus, we had proof that cleavage of this fusion protein was essentially NS3-dependent and not caused by some endogenous protease.

Secretion of the cleaved substrate was determined by assaying culture media with a commercially

SUBSTITUTE SHEET (RULE 26)

-18-

available mature IL-1 β -specific ELISA assay (R&D Systems, Minneapolis, MN). For the wild-type NS3-containing construct we detected a concentration of 2.5 μ g/ml of IL-1 β in the medium. We detected less than 0.25 μ g/ml of IL-1 β in the media of cells transfected with the mutant NS3-containing construct. Immunoprecipitation experiment utilizing the same anti-IL-1 β antibody demonstrated the presence of A4B-IL-1 β in the media of cells containing the wild type NS3-containing construct, but none from the mutant NS3-containing construct (Figure 4), thus confirming these results.

EXAMPLE 4

NS3-Specific Processing Of Mutated Pre-IL-1 β Containing An Artificial Cleavage Site And Secretion Of IL-1 β Into The Media

We confirmed that NS3 protease can cleave artificial substrates other than an HCV polypeptide by cotransfecting COS cells with the NS3-4A and either of the pre-IL-1 β -containing artificial substrate expression constructs described in Example 1C.

Co-expression of the NS3-4A and pre-IL-1 β * substrate sequences resulted in rapid cleavage of the substrate and concomitant secretion of a 19 Kd IL-1 β into the media. Secretion was quantitated using an ELISA specific for the processed form of IL-1 β . An immunoblot of cell lysates from these transformants demonstrated the presence of both cleaved and uncleaved substrate (Figure 5, NS3-4A + IL-1 β * lane). The same experiment was performed using cells that were metabolically labelled with [35 S]-methionine, followed by immunoprecipitation of the media with the processed IL-1 β -specific antibody. The results of the immunoprecipitation experiment are shown in Figure 6, NS3-4A + pre-IL-1 β * lanes.

When we coexpressed NS3-4A and the pre-IL-1 β (CSM) sequences, we also observed cleavage of the substrate at the predicted Cys₁₁₆-Ser₁₁₇ site. Both cleaved and uncleaved forms were observed in cell lysates using

-19-

immunoblotting specific for IL-1 β (Figure 5, NS3-4A + IL-1 β (CSM) lane). Immunoprecipitation of the media from [³⁵S]-methionine labelled cells also demonstrated the presence IL-1 β -containing cleavage product, but less than
5 that observed for the 5A-5B-containing pre-IL-1 β substrate (Figure 6, NS3-4A + pre-IL-1 β (CSM) lane).

EXAMPLE 5

Assay of NS3 Inhibitors

We tested the potential of compounds VH-15924
10 and VH-16075 as HCV NS3 protease inhibitors in our assays.

Transfectants expressing the NS3-4A- Δ 4B-IL-1 β were grown in the presence of varying amounts VH-15924. Even at concentrations as high as 100 μ M, we detected the
15 presence of the cleavage product, Δ 4B-IL-1 β , in the media. This indicated that VH-15924 was not an effective inhibitor of NS3 protease.

We also assayed the inhibition of cleavage and secretion of pre-IL-1 β * substrate by both VH-15924 and
20 VH-16075. VH-16075 inhibited cleavage and secretion with an IC₅₀ of 4 μ M. As in the previous experiment, VH-15924 did not completely inhibit cleavage/secretion even at concentrations of 100 μ M (Figure 7).

While I have hereinbefore presented a number of
25 embodiments of this invention, it is apparent that my basic construction can be altered to provide other embodiments which utilize the methods of this invention. Therefore, it will be appreciated that the scope of this invention is to be defined by the claims appended hereto
30 rather than the specific embodiments which have been presented hereinbefore by way of example.

-20-

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Su, Michael
- (ii) TITLE OF INVENTION: METHODS AND HOST CELLS FOR ASSAYING
EXOGENOUS AND ENDOGENOUS PROTEASE ACTIVITY
- (iii) NUMBER OF SEQUENCES: 13
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Fish & Neave
 - (B) STREET: 1251 Avenue of the Americas
 - (C) CITY: New York
 - (D) STATE: New York
 - (E) COUNTRY: United States of America
 - (F) ZIP: 10020
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Haley Jr, James F
 - (B) REGISTRATION NUMBER: 27,794
 - (C) REFERENCE/DOCKET NUMBER: VPI/95-01
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 212-596-9000
 - (B) TELEFAX: 212-596-9090

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 9401 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (ix) FEATURE:
 - (A) NAME/KEY: mat_peptide
 - (B) LOCATION: 3420..5312
 - (D) OTHER INFORMATION: /product= "NS3 protease"
- (ix) FEATURE:
 - (A) NAME/KEY: mat_peptide
 - (B) LOCATION: 5313..5474

SUBSTITUTE SHEET (RULE 26)

-21-

(D) OTHER INFORMATION: /product= "NS4A"

(ix) FEATURE:

(A) NAME/KEY: mat_peptide

(B) LOCATION: 5475..5552

(D) OTHER INFORMATION: /product= "truncated NS4B"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

| | |
|--|------|
| GCCAGCCCCC TGATGGGGGC GACACTCCAC CATAGATCAC TCCCCTGTGA GGAACACTG | 60 |
| TCTTCACGCA GAAAGCGTCT AGCCATGGCG TTAGTATGAG TGTCGTGCAG CCTCCAGGAC | 120 |
| CCCCCTCCC GGGAGAGCCA TAGTGGTCTG CGGAACCGGT GAGTACACCG GAATTGCCAG | 180 |
| GACGACCGGG TCCTTTCTTG GATAAACCCG CTCAATGCCT GGAGATTGG GCGTGCCCCC | 240 |
| GCAAGACTGC TAGCCGAGTA GTGTTGGGTC GCGAAAGGCC TTGTGGTACT GCCTGATAGG | 300 |
| GTGCTTGCGA GTGCCCCGGG AGGTCTCGTA GACCGTGCAC CATGAGCAGC AATCTAAAC | 360 |
| CTCAAAGAAA AACCAAACGT AACACCAACC GTCGCCACA GGACGTCGAG TTCCCGGGTG | 420 |
| GCGGTCAGAT CGTTGGTGA GTTTACTTGT TGCCGCGCAG GGGCCCTAGA TTGGGTGTGC | 480 |
| GCGCGACGAG GAAGACTTCC GAGCGGTCGC AACCTCGTGG TAGACGTCAG CCTATCCCCA | 540 |
| AGGCACGTCG GCCCGAGGGC AGGACCTGGG CTCAGCCCGG GTACCCTTGG CCCCTCTATG | 600 |
| GCAATGAGGG TTGCGGGTGG GCGGGATGGC TCCTGTCTCC CCGTGGCTCT CGGCCTAGCT | 660 |
| GGGGCCCCAC AGACCCCGG CGTAGGTGCG GCAATTTGGG TAAGGTCATC GATACCCTTA | 720 |
| CGTGCGGCTT CGCCGACCTC ATGGGGTACA TACCGCTCGT CGGCGCCCTT CTGAGGGCG | 780 |
| CTGCCAGGGC CCTGGCGCAT GGCCTCCGGG TTCTGGAAGA CGGCGTGAAC TATGCAACAG | 840 |
| GGAACCTTCC TGGTTGCTCT TTCTCTATCT TCCTTCTGGC CCTGCTCTCT TGCTGACTG | 900 |
| TGCCCCGCTT AGCCTACCAA GTGCGCAATT CCTCGGGGCT TTACCATGTC ACCAATGATT | 960 |
| GCCCTAATTC GAGTATTGTG TACGAGGCGG CCGATGCCAT CCTGCACACT CCGGGGTGTG | 1020 |
| TCCCTTGCGT TCGCGAGGGT AACGCCTCGA GGTGTTGGGT GGCGGTGACC CCCACGGTGG | 1080 |
| CCACCAGGGA CGGCAAACTC CCCACAACGC AGCTTCGACG TCATATCGAT CTGCTTGTGC | 1140 |
| GGAGCGCCAC CCTCTGTCA GCCCTCTACG TGGGGGACCT GTGCGGGTCT GTTTTTCTTG | 1200 |
| TTGGTCAACT GTTTACCTTC TCTCCAGGC GCCACTGGAC GACGCAAAGC TGCAATTGTT | 1260 |
| CTATCTATCC CGGCCATATA ACGGGTCATC GCATGGCATG GGATATGATG ATGAACTGGT | 1320 |
| CCCCTACGGC AGCGTTGGTG GTAGCTCAGC TGCTCCGGAT CCCACAAGCC ATCATGGACA | 1380 |
| TGATCGCTGG TGCTCACTGG GGAGTCCTGG CGGGCATAGC GTATTTCTCC ATGGTGGGGA | 1440 |
| ACTGGGCGAA GGTCTGGTA GTGCTGCTGC TATTTGCCGG CGTCGACGG GAAACCCACG | 1500 |
| TCACCGGGGG AAGTGCCGGC CACACCACGG CTGGGCTTGT TGGTCTCCTT ACACCAGGCG | 1560 |
| CCAAGCAGAA CATCCAACATG ATCAACACCA ACGGCAGTTG GCACATCAAT AGCACGGCCT | 1620 |

SUBSTITUTE SHEET (RULE 26)

-22-

| | | | | | | |
|------------|------------|------------|------------|------------|------------|------|
| TGAACTGCAA | CGATAGCCTT | ACCACCGGCT | GGTTAGCAGG | GCTCTTCTAT | CGCCACAAAT | 1680 |
| TCAACTCTTC | AGGCTGTCCT | GAGAGGTTGG | CCAGCTGCCG | ACGCCTTACC | GATTTTGCCC | 1740 |
| AGGGCTGGGG | TCCCATCAGT | TATGCCAACG | GAAGCGGCCT | TGACGAACGC | CCCTACTGTT | 1800 |
| GGCACTACCC | TCCAAGACCT | TGTGGCATTG | TGCCCCGAAA | GAGCGTGTGT | GGCCCCGTAT | 1860 |
| ATTGCTTCAC | TCCCAGCCCC | GTGGTGGTGG | GAACGACCGA | CAGGTCGGGC | GCGCCTACCT | 1920 |
| ACAGCTGGGG | TGCAAATGAT | ACGGATGTCT | TCGTCTTAA | CAACACCAGG | CCACCGCTGG | 1980 |
| GCAATTGGTT | CGGTTGTACC | TGGATGAACT | CAACTGGATT | CACCAAAGTG | TGCGGAGCGC | 2040 |
| CCCCTTGTGT | CATCGGAGGG | GTGGGCAACA | ACACCTTGCT | CTGCCCCACT | GATTGCTTCC | 2100 |
| GCAAACATCC | GGAAGCCACA | TACTCTCGGT | GCGGCTCCGG | TCCCTGGATT | ACACCCAGGT | 2160 |
| GCATGGTCGA | CTACCCGTAT | AGGCTTTGGC | ACTATCCTTG | TACTATCAAT | TACACCATAT | 2220 |
| TCAAAGTCAG | GATGTACGTG | GGAGGGGTCG | AGCACAGGCT | GGAAGCGGCC | TGCAACTGGA | 2280 |
| CGCGGGGCGA | ACGCTGTGAT | CTGGAAGACA | GGGACAGGTC | CGAGCTCAGC | CCATTGCTGC | 2340 |
| TGTCCACCAC | ACAGTGGCAG | GTCTTCCGT | GTTCTTTCAC | GACCCTGCCA | GCCTTGTCCA | 2400 |
| CCGGCCTCAT | CCACCTCCAC | CAGAACATTG | TGGACGTGCA | GTACTTGTAC | GGGGTGGGGT | 2460 |
| CAAGCATCGC | GTCCTGGGCC | ATTAAGTGGG | AGTACGTCGT | TCTCCTGTTC | CTTCTGCTTG | 2520 |
| CAGACGCGCG | CGTCTGCTCC | TGCTTGTGGA | TGATGTTACT | CATATCCCAA | GCGGAGGCGG | 2580 |
| CTTTGGAGAA | CCTCGTAATA | CTCAATGCAG | CATCCCTGGC | CGGGACGCAC | GGTCTTGTGT | 2640 |
| CCTTCCTCGT | GTTCTTCTGC | TTTGCCTGGT | ATCTGAAGGG | TAGGTGGGTG | CCCGGAGCGG | 2700 |
| TCTACGCCTT | CTACGGGATG | TGGCCTCTCC | TCCTGCTCCT | GCTGGCGTTG | CCTCAGCGGG | 2760 |
| CATACGCACT | GGACACGGAG | GTGGCCGCGT | CGTGTGGCGG | CGTTGTTCTT | GTCGGGTAA | 2820 |
| TGGCGCTGAC | TCTGTCACCA | TATTACAAGC | GCTATATCAG | CTGGTGCATG | TGGTGGCTTC | 2880 |
| AGTATTTTCT | GACCAGAGTA | GAAGCGCAAC | TGCACGTGTG | GGTTCCCCC | CTCAACGTCC | 2940 |
| GGGGGGGGCG | CGATGCCGTC | ATCTTACTCA | TGTGTGTTGT | ACACCCGACT | CTGGTATTTG | 3000 |
| ACATCACCAA | ACTACTCCTG | GCCATCTTCG | GACCCCTTTG | GATTCTTCAA | GCCAGTTTGC | 3060 |
| TTAAAGTCCC | CTACTTCGTG | CGCGTTCAAG | GCCTTCTCCG | GATCTGCGCG | CTAGCGCGGA | 3120 |
| AGATAGCCGG | AGGTCATTAC | GTGCAAATGG | CCATCATCAA | GTGGGGGCGG | CTTACTGGCA | 3180 |
| CCTATGTGTA | TAACCATCTC | ACCCCTCTTC | GAGACTGGGC | GCACAACGGC | CTGCGAGATC | 3240 |
| TGGCCGTGGC | TGTGGAACCA | GTCTTCTTCT | CCCGAATGGA | GACCAAGCTC | ATCACGTGGG | 3300 |
| GGGCAGATAC | CGCCGCGTGC | GGTGACATCA | TCAACGGCTT | GCCCGTCTCT | GCCCGTAGGG | 3360 |
| GCCAGGAGAT | ACTGCTTGGA | CCAGCCGACG | GAATGGTCTC | CAAGGGGTGG | AGGTTGCTGG | 3420 |
| CGCCCATCAC | GGCGTACGCC | CAGCAGACGA | GAGGCCTCCT | AGGGTGTATA | ATCACCAGCC | 3480 |
| TGACTGGCCG | GGACAAAAC | CAAGTGGAGG | GTGAGGTCCA | GATCGTGTCA | ACTGCTACCC | 3540 |

SUBSTITUTE SHEET (RULE 26)

-23-

| | |
|--|------|
| AAACCTTCCT GGCAACGTGC ATCAATGGGG TATGCTGGAC TGTCTACCAC GGGGCCGGAA | 3600 |
| CGAGGACCAT CGCATCACCC AAGGGTCCTG TCATCCAGAT GTATACCAAT GTGGACCAAG | 3660 |
| ACCTTGTGGG CTGGCCCGCT CCTCAAGGTT CCCGCTCATT GACACCCTGC ACCTGCGGCT | 3720 |
| CCTCGGACCT TTACCTGGTT ACGAGGCACG CCGACGTCAT TCCCCTGCGC CGGCGAGGTG | 3780 |
| ATAGCAGGGG TAGCCTGCTT TCGCCCCGGC CCATTTCTTA CCTAAAAGGC TCCTCGGGGG | 3840 |
| GTCCGCTGTT GTGCCCCGCG GGACACGCGG TGGGCTATT CAGGGCCGCG GTGTGCACCC | 3900 |
| GTGGAGTGAC CAAGGCGGTG GACTTTATCC CTGTGGAGAA CCTAGAGACA ACCATGAGAT | 3960 |
| CCCCGGTGT CACGGACAAC TCCTCTCCAC CAGCAGTGCC CCAGAGCTTC CAGGTGGCCC | 4020 |
| ACCTGCATGC TCCCACCGGC AGTGGTAAGA GCACCAAGGT CCCGGCTGCG TACGCAGCCC | 4080 |
| AGGGGTACAA GGTGTTGGTG CTCAACCCCT CTGTTGCTGC AACGCTGGGC TTTGGTGCTT | 4140 |
| ACATGTCCAA GGCCCATGGG GTCGATCCTA ATATCAGGAC CGGGGTGAGA ACAATTACCA | 4200 |
| CTGGCAGCCC CATCAGTAC TCCACCTACG GCAAGTTCCT TGCCGACGGC GGGTGCTCAG | 4260 |
| GAGGCGCTTA TGACATAATA ATTTGTGACG AGTGCCACTC CACGGATGCC ACATCCATCT | 4320 |
| TGGGCATCGG CACTGTCCTT GACCAAGCAG AGACTGCGGG GGCGAGATTG GTTGTGCTCG | 4380 |
| CCACTGCTAC CCCTCCGGGC TCCGTCACTG TGTCCCATCC TAACATCGAG GAGGTTGCTC | 4440 |
| TGTCCACCAC CGGAGAGATC CTTTCTACG GCAAGGCTAT CCCCTCGAG GTGATCAAGG | 4500 |
| GGGGAAGACA TCTCATCTTC TGTCACCTCA AGAAGAAGTG CGACGAGCTC GCCCGGAAGC | 4560 |
| TGGTCGCATT GGGCATCAAT GCCGTGGCCT ACTACCGCGG ACTTGACGTG TCTGTCATCC | 4620 |
| CGACCAACGG CGATGTTGTC GTCGTGTCGA CCGATGCTCT CATGACTGGC TTTACCGGGC | 4680 |
| ACTTCGACTC TGTGATAGAC TGCAACACGT GTGTCACTCA GACAGTCGAT TTCAGCCTTG | 4740 |
| ACCCTACCTT TACCATTGAG ACAACCACGC TCCCCCAGGA TGCTGTCTCC AGGACTCAGC | 4800 |
| GCCGGGGCAG GACTGGCAGG GGGGAAGCCAG GCATCTACAG ATTTGTGGCA CCGGGGGAGC | 4860 |
| GCCCCCTCCG CATGTTGAC TCGTCCGTCC TCTGTGAGTG CTATGACGCG GGCTGTGCTT | 4920 |
| GGTATGAGCT CATGCCCCGC GAGACTACAG TTAGGCTACG AGCGTACATG AACACCCCGG | 4980 |
| GGCTTCCCGT GTGCCAGGAC CATCTTGAAT TTTGGGAGGG CGTCTTTACG GGCCTCACCC | 5040 |
| ATATAGATGC CCACTTTCTA TCCCAGACAA AGCAGAGTGG GGAGAACTTT CCTTACCTGG | 5100 |
| TAGCGTACCA AGCCACCGTG TGCCTAGGG CTCAAGCCCC TCCCCCATCG TGGGACCAGA | 5160 |
| TGTGGAAGTG TTTGATCCGC CTTAAACCCA CCCTCCATGG GCCAACACCC CTGCTATACA | 5220 |
| GACTGGGCGC TGTTGAGAAT GAAGTCACCC TGACGCACCC AATCACCAA TACATCATGA | 5280 |
| CATGCATGTC GGCCGACCTG GAGGTCGTCA CGAGCACCTG GGTGCTCGTT GCGGGCGTCC | 5340 |
| TGGCTGCTCT GGCCGCGTAT TGCCTGTCAA CAGGCTGCGT GGTGATAGTG GGCAGGATTG | 5400 |
| TCTTGTCGGG GAAGCCGGCA ATTATACCTG ACAGGGAGGT TCTCTACCAG GAGTTCGATG | 5460 |

SUBSTITUTE SHEET (RULE 26)

-24-

| | |
|---|------|
| AGATGGAAGA GTGCTCTCAG CACTTACCGT ACATCGAGCA AGGGATGATG CTCGCTGAGC | 5520 |
| AGTTCAAGCA GAAGGCCCTC GGCCTCCTGC AGACCGCGTC CCGCCATGCA GAGGTTATCA | 5580 |
| CCCCTGCTGT CCAGACCAAC TGGCAGAAAC TCGAGGTCTT CTGGGCGAAG CACATGTGGA | 5640 |
| ATTTTCATCAG TGGGATACAA TATTTGGCGG GCCTGTCAAC GCTGCCTGGT AACCCCGCCA | 5700 |
| TTGCTTCATT GATGGCTTTT ACAGCTGCCG TCACCAGCCC ACTAACCCTT GGCCAAACCC | 5760 |
| TCCTCTTCAA CATATTGGGG GGGTGGGTGG CTGCCCAGCT CGCCGCCCCC GGTGCCGCTA | 5820 |
| CCGCCTTTGT GGGCGCTGGC TTAGCTGGCG CCGCCATCGG CAGCGTTGGA CTGGGGAAGG | 5880 |
| TCCTCGTGGA CATTCTTGCA GGGTATGGCG CGGGCGTGGC GGGAGCTCTT GTAGCATTCA | 5940 |
| AGATCATGAG CGGTGAGGTC CCCTCCACGG AGGACCTGGT CAATCTGCTG CCCGCCATCC | 6000 |
| TCTCGCCTGG AGCCCTTGTA GTCGGTGTGG TCTGCGCAGC AATACTGCGC CGGCACGTTG | 6060 |
| GCCCGGGCGA GGGGGCAGTG CAATGGATGA ACCGGCTAAT AGCCTTCGCC TCCCGGGGGA | 6120 |
| ACCATGTTTC CCCCACGCAC TACGTGCCGG AGAGCGATGC AGCCGCCCCG GTCACTGCCA | 6180 |
| TACTCAGCAG CCTCACTGTA ACCCAGCTCC TGAGGCGACT ACATCAGTGG ATAAGCTCGG | 6240 |
| AGTGTACCAC TCCATGCTCC GGCTCCTGGC TAAGGGACAT CTGGGACTGG ATATGCGAGG | 6300 |
| TGCTGAGCGA CTTTAAGACC TGGCTGAAAG CCAAGCTCAT GCCACAAC TG CCTGGGATTC | 6360 |
| CCTTTGTGTC CTGCCAGCGC GGGTATAGGG GGGTCTGGCG AGGAGACGGC ATTATGCACA | 6420 |
| CTCGCTGCCA CTGTGGAGCT GAGATCACTG GACATGTCAA AAACGGGACG ATGAGGATCG | 6480 |
| TCGGTCCTAG GACCTGCAGG AACATGTGGA GTGGGACGTT CCCCATTAAC GCCTACACCA | 6540 |
| CGGGCCCCCTG TACTCCCCCTT CCTGCGCCGA ACTATAAGTT CGCGCTGTGG AGGGTGTCTG | 6600 |
| CAGAGGAATA CGTGGAGATA AGGCGGGTGG GGGACTTCCA CTACGTATCG GGTATGACTA | 6660 |
| CTGACAATCT TAAATGCCCG TGCCAGATCC CATCGCCCGA ATTTTTCACA GAATTGGACG | 6720 |
| GGGTGCGCCT ACATAGGTTT GCGCCCCCTT GCAAGCCCTT GCTGCGGGAG GAGGTATCAT | 6780 |
| TCAGAGTAGG ACTCCACGAG TACCCGGTGG GGTGCAATT ACCTGCGAG CCCGAACCGG | 6840 |
| ACGTAGCCGT GTTGACGTCC ATGCTCACTG ATCCCTCCA TATAACAGCA GAGGCGGCCG | 6900 |
| GGAGAAGGTT GCGGAGAGGG TCACCCCTT CTATGGCCAG CTCCTCGGCC AGCCAGCTGT | 6960 |
| CCGCTCCATC TCTCAAGCA ACTTGACCG CCAACCATGA CTCCCTGAC GCCGAGCTCA | 7020 |
| TAGAGGCTAA CCTCCTGTGG AGGCAGGAGA TGGGCGGCAA CATCACCAGG GTTGAGTCAG | 7080 |
| AGAACAAGT GGTGATTCTG GACTCCTTCG ATCCGCTTGT GGCAGAGGAG GATGAGCGGG | 7140 |
| AGGTCTCCGT ACCCGCAGAA ATTCTGCGGA AGTCTCGGAG ATTGCCCCGG GCCCTGCCCC | 7200 |
| TTTGGGCGCG GCCGGACTAC AACCCCCCG TAGTAGAGAC GTGGAAAAAG CCTGACTACG | 7260 |
| AACCACCTGT GGTCCATGGC TGCCCGCTAC CACCTCCAG GTCCCTCCT GTGCCTCCGC | 7320 |
| CTCGGAAAAA GCGTACGGTG GTCCTCACCG AATCAACCCT ACCTACTGCC TTGGCCGAGC | 7380 |

SUBSTITUTE SHEET (RULE 26)

| | |
|--|------|
| TTGCCACCAA AAGTTTGGC AGCTCCTCAA CTTCCGGCAT TACGGGCGAC AATATGACAA | 7440 |
| CATCCTCTGA GCGCGCCCT TCTGGCTGCC CCCCCGACTC CGACGTTGAG TCCTATTCTT | 7500 |
| CCATGCCCCC CTTGGAGGGG GAGCCTGGGG ATCCGGATTT CAGCGACGGG TCATGGTCGA | 7560 |
| CGGTCAAGTAG TGGGGCCGAC ACGGAAGATG TCGTGTGCTG CTCAATGTCT TATACCTGGA | 7620 |
| CAGGCGCACT CGTCACCCCG TCGCTGCGG AAGAACAAA ACTGCCCCATC AACGCACTGA | 7680 |
| GCAACTCGTT GCTACGCCAT CACAATCTGG TATATTCCAC CACTTCACGC AGTGCTTGCC | 7740 |
| AAAGGCAGAA GAAAGTCACA TTTGACAGAC TGCAAGTTCT GGACAGCCAT TACCAGGACG | 7800 |
| TGCTCAAGGA GGTCAAAGCA GCGGCGTCAA AAGTGAAGGC TAACCTGCTA TCCGTAGAGG | 7860 |
| AAGCTTGCGC CTTGACGCCC CCACATTCAG CCAAATCCAA GTTTGGCTAT GGGGCAAAAG | 7920 |
| ACGTCCGTTG CCATGCCAGA AAGGCCGTAG CCCACATCAA CTCCGTGTGG AAAGACCTTC | 7980 |
| TGGAAGACAG TGTAACACCA ATAGACACTA TCATCATGGC CAAGAACGAG GTCTTCTGCG | 8040 |
| TTGAGCTGA GAAGGGGGGT CGTAAGCCAG CTCGTCTCAT CGTGTTCCTC GACCTGGGCG | 8100 |
| TGCGCGTGTG CGAGAAGATG GCCCTGTACG ACGTGGTTAG CAAACTCCCC CTGGCCGTGA | 8160 |
| TGGGAAGCTC CTACGGATTC CAATACTCAC CAGGACAGCG GGTGAATTC CTCGTGCAAG | 8220 |
| CGTGAAGTC CAAGAAGACC CCGATGGGGT TCCCGTATGA TACCCGCTGT TTTGACTCCA | 8280 |
| CAGTCACTGA GAGCGACATC CGTACGGAGG AGGCAATTTA CCAATGTTGT GACCTGGACC | 8340 |
| CCCAAGCCCC CGTGGCCATC AAGTCCCTCA CTGAGAGGCT TTATGTTGGG GGCCCTCTTA | 8400 |
| CCAATTCAAG GGGGAAAAC TCGGCTATC GCAGGTGCCG CGCGAGCGGC GTACTGACAA | 8460 |
| CTAGCTGTGG TAACACCCTC ACTTGCTACA TCAAGGCCCG GGCAGCCCGT CGAGCCGCAG | 8520 |
| GGCTCCAGGA CTGCACCATG CTCGTGTGTG GCGACGACTT AGTCGTTATC TGTGAAAGTG | 8580 |
| CGGGGGTCCA GGAGGACGCG GCGAGCCTGA GAGCCTTTAC GGAGGCTATG ACCAGGTACT | 8640 |
| CGCCCCCCCC CGGGGACCCC CCACAACCAG AATACGACTT GGAGCTTATA ACATCATGCT | 8700 |
| CCTCCAACGT GTCAGTCGCC CACGACGGCG CTGGA AAAAG GGTCTACTAC CTTACCCGTG | 8760 |
| ACCCTACAAC CCCCCTCGCG AGAGCCGCGT GGGAGACAGC AAGACACACT CCAGTCAATT | 8820 |
| CCTGGCTAGG CAACATAATC ATGTTTGCCC CCACACTGTG GGCGAGGATG ATACTGATGA | 8880 |
| CCCATTCTT TAGCGTCCTC ATAGCCAGGG ATCAGCTTGA ACAGGCTCTT AACTGTGAGA | 8940 |
| TCTACGCAGC CTGCTACTCC ATAGAACCAC TGGATCTACC TCCAATCATT CAAAGACTCC | 9000 |
| ATGGCCTCAG CGCATTTTTA CTCCACAGTT ACTCTCCAGG TGAAGTCAAT AGGGTGGCCG | 9060 |
| CATGCCTCAG AAAACTTGGG GTCCCGCCCT TGCGAGCTTG GAGACACCGG GCGCGAGCG | 9120 |
| TCCGCGCTAG GCTTCTGTCC AGGGGAGGCA GGGCTGCCAT ATGTGGCAAG TACCTCTTCA | 9180 |
| ACTGGGCAGT AAGAACAAAG CTCAAATCA CTCCAATAGC GGCCGCTGGC CGGCTGGACT | 9240 |
| TGTCCGTTG GTTACGGCT GGCTACAGCG GGGGAGACAT TTATCACAGC GTGTCTCATG | 9300 |

-26-

CCCGGCCCCG CTGGTTCTGG TTTTGCCTAC TCCTGCTCGC TGCAGGGGTA GGCATCTACC 9360
TCCTCCCCAA CCGGTGAACG GGGAGCTAGA CACTCCGGCC T 9401

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 47 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "oligonucleotide primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GGACTAGTCT GCAGTCTAGA GCTCCATGGC GCCCATCAGC GCGTACG 47

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 44 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "oligonucleotide primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CGGCTGCACC TCCAGCAGTG CATTTTAGAT CTTAAGTCTA GAAG 44

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 44 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "oligonucleotide primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AAGCTACTCT ACCTTCTCAC GATTTTAGAT CTTAAGTCTA GAAG 44

(2) INFORMATION FOR SEQ ID NO:5:

SUBSTITUTE SHEET (RULE 26)

-27-

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 45 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "oligonucleotide primer"
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GGATCTCTGT TGGTACTCTA GGATTTTAGA TCTTAAGTCT AGAAG

45

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 64 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "OLIGONUCLEOTIDE DUPLEX"
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (ix) FEATURE:
 - (A) NAME/KEY: misc_feature
 - (B) LOCATION: 1..4
 - (D) OTHER INFORMATION: /product= "SINGLE STRANDED REGION ON CODING STRAND"
- (ix) FEATURE:
 - (A) NAME/KEY: misc_feature
 - (B) LOCATION: 61..64
 - (D) OTHER INFORMATION: /product= "SINGLE STRANDED REGION ON COMPLEMENTARY STRAND"
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TGCACGGCGC CGACACGGAA GATGTCGTGT GCTGCTCAAT GTCTTATACC TGGACAGGCG
TGCA

60

64

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (iii) HYPOTHETICAL: NO

SUBSTITUTE SHEET (RULE 26)

-28-

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Gly | Ala | Asp | Thr | Glu | Asp | Val | Val | Cys | Cys | Ser | Met | Ser | Tyr | Thr | Trp |
| 1 | | | | 5 | | | | 10 | | | | | 15 | | |
| | | | | | | | | | | | | | | | |
| Thr | Gly | Val | His | | | | | | | | | | | | |
| | | | 20 | | | | | | | | | | | | |

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "oligonucleotide primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GGACTAGTCT GCAGTCTAGA GCTCCATGGC GCCCATCAGC GCGTACG

47

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "oligonucleotide primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CGGGAGCCGG AGGACGTCTG GCGCAGG

27

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1497 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

SUBSTITUTE SHEET (RULE 26)

-29-

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 87..893

(ix) FEATURE:

(A) NAME/KEY: misc_feature
(B) LOCATION: 426..427
(D) OTHER INFORMATION: /label= ApaLisite

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

| | |
|---|-----|
| ACCAACCTCT TCGAGGCACA AGGCACAACA GGCTGCTCTG GGATTCTCTT CAGCCAATCT | 60 |
| TCATTGCTCA AGTGTCTGAA GCAGCC ATG GCA GAA GTA CCT GAG CTC GCC AGT | 113 |
| Met Ala Glu Val Pro Glu Leu Ala Ser | |
| 1 5 | |
| GAA ATG ATG GCT TAT TAC AGT GGC AAT GAG GAT GAC TTG TTC TTT GAA | 161 |
| Glu Met Met Ala Tyr Tyr Ser Gly Asn Glu Asp Asp Leu Phe Phe Glu | |
| 10 15 20 25 | |
| GCT GAT GGC CCT AAA CAG ATG AAG TGC TCC TTC CAG GAC CTG GAC CTC | 209 |
| Ala Asp Gly Pro Lys Gln Met Lys Cys Ser Phe Gln Asp Leu Asp Leu | |
| 30 35 40 | |
| TGC CCT CTG GAT GGC GGC ATC CAG CTA CGA ATC TCC GAC CAC CAC TAC | 257 |
| Cys Pro Leu Asp Gly Gly Ile Gln Leu Arg Ile Ser Asp His His Tyr | |
| 45 50 55 | |
| AGC AAG GGC TTC AGG CAG GCC GCG TCA GTT GTT GTG GCC ATG GAC AAG | 305 |
| Ser Lys Gly Phe Arg Gln Ala Ala Ser Val Val Val Ala Met Asp Lys | |
| 60 65 70 | |
| CTG AGG AAG ATG CTG GTT CCC TGC CCA CAG ACC TTC CAG GAG AAT GAC | 353 |
| Leu Arg Lys Met Leu Val Pro Cys Pro Gln Thr Phe Gln Glu Asn Asp | |
| 75 80 85 | |
| CTG AGC ACC TTC TTT CCC TTC ATC TTT GAA GAA GAA CCT ATC TTC TTC | 401 |
| Leu Ser Thr Phe Phe Pro Phe Ile Phe Glu Glu Glu Pro Ile Phe Phe | |
| 90 95 100 105 | |
| GAC ACA TGG GAT AAC GAG GCT TAT GTG CAC GAT GCA CCT GTA CGA TCA | 449 |
| Asp Thr Trp Asp Asn Glu Ala Tyr Val His Asp Ala Pro Val Arg Ser | |
| 110 115 120 | |
| CTG AAC TGC ACG CTC CGG GAC TCA CAG CAA AAA AGC TTG GTG ATG TCT | 497 |
| Leu Asn Cys Thr Leu Arg Asp Ser Gln Gln Lys Ser Leu Val Met Ser | |
| 125 130 135 | |
| GGT CCA TAT GAA CTG AAA GCT CTC CAC CTC CAG GGA CAG GAT ATG GAG | 545 |
| Gly Pro Tyr Glu Leu Lys Ala Leu His Leu Gln Gly Gln Asp Met Glu | |
| 140 145 150 | |
| CAA CAA GTG GTG TTC TCC ATG TCC TTT GTA CAA GGA GAA GAA AGT AAT | 593 |
| Gln Gln Val Val Phe Ser Met Ser Phe Val Gln Gly Glu Glu Ser Asn | |
| 155 160 165 | |
| GAC AAA ATA CCT GTG GCC TTG GGC CTC AAG GAA AAG AAT CTG TAC CTG | 641 |
| Asp Lys Ile Pro Val Ala Leu Gly Leu Lys Glu Lys Asn Leu Tyr Leu | |
| 170 175 180 185 | |
| TCC TGC GTG TTG AAA GAT GAT AAG CCC ACT CTA CAG CTG GAG AGT GTA | 689 |

SUBSTITUTE SHEET (RULE 26)

-30-

| | |
|--|------|
| Ser Cys Val Leu Lys Asp Asp Lys Pro Thr Leu Gln Leu Glu Ser Val | |
| 190 195 200 | |
| GAT CCC AAA AAT TAC CCA AAG AAG AAG ATG GAA AAG CGA TTT GTC TTC | 737 |
| Asp Pro Lys Asn Tyr Pro Lys Lys Lys Met Glu Lys Arg Phe Val Phe | |
| 205 210 215 | |
| AAC AAG ATA GAA ATC AAT AAC AAG CTG GAA TTT GAG TCT GCC CAG TTC | 785 |
| Asn Lys Ile Glu Ile Asn Asn Lys Leu Glu Phe Glu Ser Ala Gln Phe | |
| 220 225 230 | |
| CCC AAC TGG TAC ATC AGC ACC TCT CAA GCA GAA AAC ATG CCC GTC TTC | 833 |
| Pro Asn Trp Tyr Ile Ser Thr Ser Gln Ala Glu Asn Met Pro Val Phe | |
| 235 240 245 | |
| CTG GGA GGG ACC AAA GGC GGC CAG GAT ATA ACT GAC TTC ACC ATG CAA | 881 |
| Leu Gly Gly Thr Lys Gly Gly Gln Asp Ile Thr Asp Phe Thr Met Gln | |
| 250 255 260 265 | |
| TTT GTG TCT TCC TAAAGAGAGC TGTACCCAGA GAGTCCTGTG CTGAATGTGG | 933 |
| Phe Val Ser Ser | |
| ACTCAATCCC TAGGGCTGGC AGAAAGGGAA CAGAAAGGTT TTTGAGTACG GCTATAGCCT | 993 |
| GGACTTTCCT GTTGTCTACA CCAATGCCCA ACTGCCTGCC TTAGGGTAGT GCTAAGAGGA | 1053 |
| TCTCCTGTCC ATCAGCCAGG ACAGTCAGCT CTCTCCTTTC AGGGCCAATC CCCAGCCCTT | 1113 |
| TTGTTGAGCC AGGCCTCTCT CACCTCTCCT ACTCACTTAA AGCCCGCCTG ACAGAAACCA | 1173 |
| CGGCCACATT TGGTTCTAAG AAACCCTCTG TCATTGCTC CCACATTCTG ATGAGCAACC | 1233 |
| GCTTCCCTAT TTATTTATTT ATTTGTTTGT TTGTTTTATT CATTGGTCTA ATTTATTCAA | 1293 |
| AGGGGGCAAG AAGTAGCAGT GTCTGTAAAA GAGCCTAGTT TTTAATAGCT ATGGAATCAA | 1353 |
| TTCAATTTGG ACTGGTGTGC TCTCTTTAAA TCAAGTCCTT TAATTAAGAC TGAAAATATA | 1413 |
| TAAGCTCAGA TTATTTAAAT GGGAAATATTT ATAAATGAGC AAATATCATA CTGTTCAATG | 1473 |
| GTCTGAAAT AAACCTCTCT GAAG | 1497 |

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 269 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

| |
|---|
| Met Ala Glu Val Pro Glu Leu Ala Ser Glu Met Met Ala Tyr Tyr Ser |
| 1 5 10 15 |
| Gly Asn Glu Asp Asp Leu Phe Phe Glu Ala Asp Gly Pro Lys Gln Met |
| 20 25 30 |
| Lys Cys Ser Phe Gln Asp Leu Asp Leu Cys Pro Leu Asp Gly Gly Ile |
| 35 40 45 |

SUBSTITUTE SHEET (RULE 26)

-31-

Gln Leu Arg Ile Ser Asp His His Tyr Ser Lys Gly Phe Arg Gln Ala
 50 55 60
 Ala Ser Val Val Val Ala Met Asp Lys Leu Arg Lys Met Leu Val Pro
 65 70 75 80
 Cys Pro Gln Thr Phe Gln Glu Asn Asp Leu Ser Thr Phe Phe Pro Phe
 85 90 95
 Ile Phe Glu Glu Glu Pro Ile Phe Phe Asp Thr Trp Asp Asn Glu Ala
 100 105 110
 Tyr Val His Asp Ala Pro Val Arg Ser Leu Asn Cys Thr Leu Arg Asp
 115 120 125
 Ser Gln Gln Lys Ser Leu Val Met Ser Gly Pro Tyr Glu Leu Lys Ala
 130 135 140
 Leu His Leu Gln Gly Gln Asp Met Glu Gln Gln Val Val Phe Ser Met
 145 150 155 160
 Ser Phe Val Gln Gly Glu Glu Ser Asn Asp Lys Ile Pro Val Ala Leu
 165 170 175
 Gly Leu Lys Glu Lys Asn Leu Tyr Leu Ser Cys Val Leu Lys Asp Asp
 180 185 190
 Lys Pro Thr Leu Gln Leu Glu Ser Val Asp Pro Lys Asn Tyr Pro Lys
 195 200 205
 Lys Lys Met Glu Lys Arg Phe Val Phe Asn Lys Ile Glu Ile Asn Asn
 210 215 220
 Lys Leu Glu Phe Glu Ser Ala Gln Phe Pro Asn Trp Tyr Ile Ser Thr
 225 230 235 240
 Ser Gln Ala Glu Asn Met Pro Val Phe Leu Gly Gly Thr Lys Gly Gly
 245 250 255
 Gln Asp Ile Thr Asp Phe Thr Met Gln Phe Val Ser Ser
 260 265

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "oligonucleotide primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CTCGGCCTCC TGCAGGCACC TGTACGATCA CTGAAC

36

(2) INFORMATION FOR SEQ ID NO:13:

SUBSTITUTE SHEET (RULE 26)

-31/1-

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 32 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "oligonucleotide primer"
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GTAAACACA GAAGGATTTT AGATCTTAAG GG

32

-32-

CLAIMS

I claim:

1. A method for assaying exogenous protease activity in a host cell comprising the steps of:

(a) incubating a host cell transformed with a first nucleotide sequence encoding an exogenous protease and a second nucleotide sequence encoding an artificial polypeptide substrate;

wherein said substrate comprises:

(i) a cleavage site for said exogenous protease; and

(ii) a polypeptide that is secreted out of said cell following cleavage by said exogenous protease; under conditions which cause said exogenous protease and said artificial substrate to be expressed;

(b) separating said host cell from its growth media under non-lytic conditions; and

(c) assaying said growth media for the presence of said secreted polypeptide.

2. A method for assaying endogenous protease activity in a host cell comprising the steps of:

(a) incubating a host cell transformed with a nucleotide sequence encoding an artificial polypeptide substrate;

wherein said substrate comprises:

(i) a cleavage site for said endogenous protease; and

(ii) a polypeptide that is secreted out of said cell following cleavage by said endogenous protease; under conditions which cause said artificial substrate to be expressed;

(b) separating said host cell from its growth media under non-lytic conditions; and

(c) assaying said growth media for the presence of said secreted polypeptide.

-33-

3. A method for identifying a compound as an inhibitor of a protease comprising the steps of:

(a) assaying the activity of a protease in the absence of said compound by a method according to claim 1 or 2;

(b) assaying the activity of a protease in the presence of said compound by a method according to claim 1 or 2, wherein said compound is added to the host cells during said incubation of said host cells; and

(c) comparing the results of step (a) with the results of step (b).

4. The method according to claim 1 or claim 3, insofar as it depends from claim 1, wherein said first nucleotide sequence and said second nucleotide sequence encode a single polypeptide.

5. The method according to claim 4, wherein said first and second nucleotide sequences encode NS3-4A-Δ4B-IL-1β.

6. The method according to any one of claims 1 to 3, wherein said first nucleotide sequence encodes a viral protease or an enzymatically active fragment thereof.

7. The method according to claim 6, wherein said first nucleotide sequence encodes hepatitis C virus NS3 protease, an NS3-4A fusion protein or amino acids 1-180 of NS3 protease.

8. The method according to any one of claims 1 to 3, wherein said secreted polypeptide is selected from polypeptides comprising mature IL-1β, mature IL-1α, basic fibroblast growth factor and endothelial-monocyte activating polypeptide II.

9. The method according to claim 8, wherein said secreted polypeptide comprises mature IL-1β.

SUBSTITUTE SHEET (RULE 26)

-34-

10. The method according to claim 9, wherein said artificial polypeptide substrate is selected from pre-IL-1 β * or pre-IL-1 β (CSM).

11. A host cell transformed with a nucleotide sequence encoding an artificial polypeptide substrate, wherein said substrate comprises:

(a) a cleavage site for said exogenous protease; and

(b) a polypeptide that is secreted out of said cell following cleavage by said exogenous protease; said host cell being capable of expressing said protease and said substrate.

12. A host cell transformed with a first nucleotide sequence encoding an exogenous protease and a second nucleotide sequence encoding an artificial polypeptide substrate, wherein said substrate comprises:

(a) a cleavage site for said exogenous protease; and

(b) a polypeptide that is secreted out of said cell following cleavage by said exogenous protease; said host cell being capable of expressing said protease and said substrate.

13. The host cell according to claim 11 or 12, wherein said secreted polypeptide is selected from polypeptides comprising mature IL-1 β , mature IL-1 α , basic fibroblast growth factor and endothelial-monocyte activating polypeptide II.

14. The host cell according to claim 13, wherein said secreted polypeptide comprises mature IL-1 β .

SUBSTITUTE SHEET (RULE 26)

-35-

15. The host cell according to claim 14, wherein said artificial polypeptide substrate is selected from pre-IL-1 β * or pre-IL-1 β (CSM).

16. The host cell according to claim 12, wherein said first nucleotide sequence and said second nucleotide sequence encode a single polypeptide.

17. The host cell according to claim 16, wherein said first and second nucleotide sequences encode NS3-4A- Δ 4B-IL-1 β .

18. The host cell according to claim 12, wherein said first nucleotide sequence encodes a viral protease or an enzymatically active fragment thereof.

19. The host cell according to claim 18, wherein said first nucleotide sequence encodes hepatitis C virus NS3 protease, an NS3-4A fusion protein or amino acids 1-180 of NS3 protease.

20. The host cell according to claim 11 or 12, selected from E. coli, Bacillus, other bacteria, yeast and other fungi, plant cells, insect cells, mammalian cells.

21. The host cell according to claim 20, wherein said host cell is a mammalian cell.

22. The host cell according to claim 21, wherein said host cell is a COS cell.

23. A recombinant DNA molecule comprising a DNA sequence encoding an artificial substrate selected from pre-IL-1 β * and pre-IL-1 β (CSM).

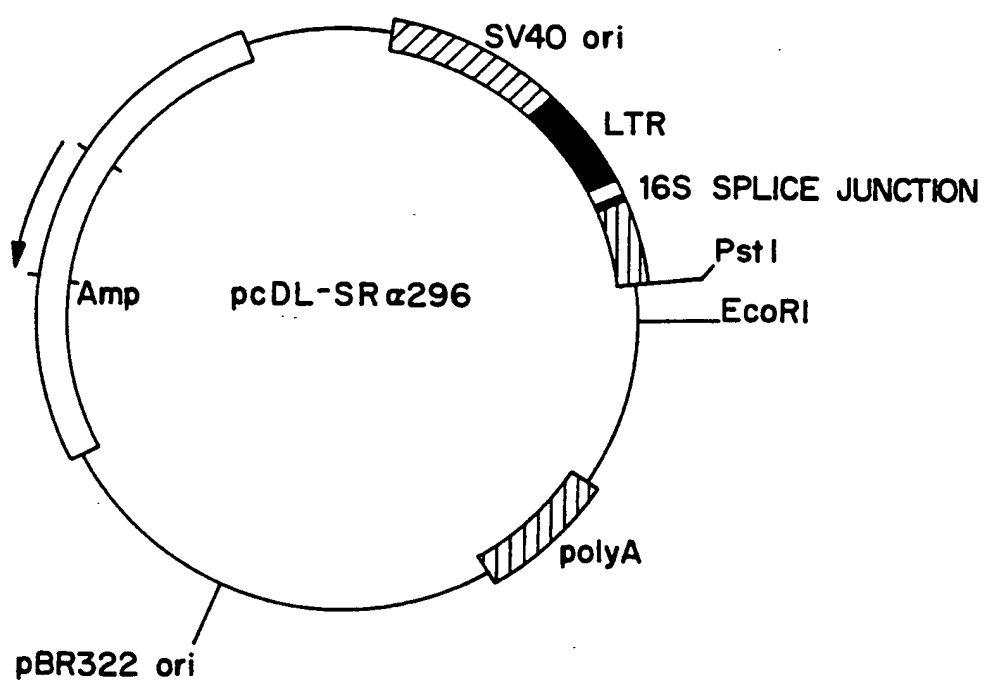


FIG. 1

2 / 7

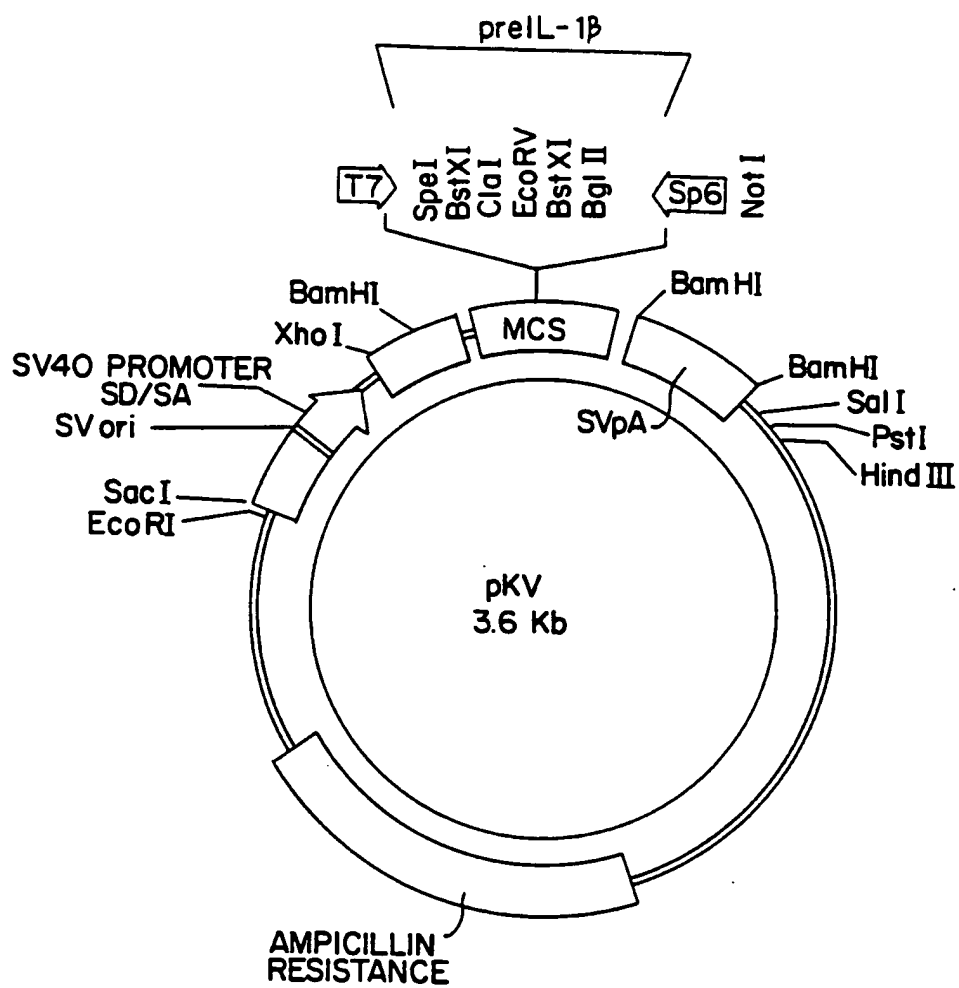


FIG. 2

3 / 7

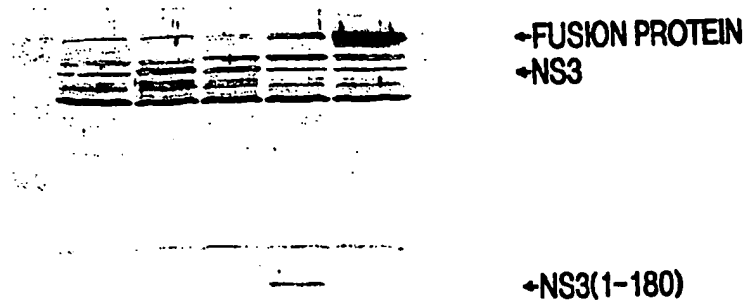


FIG. 3A

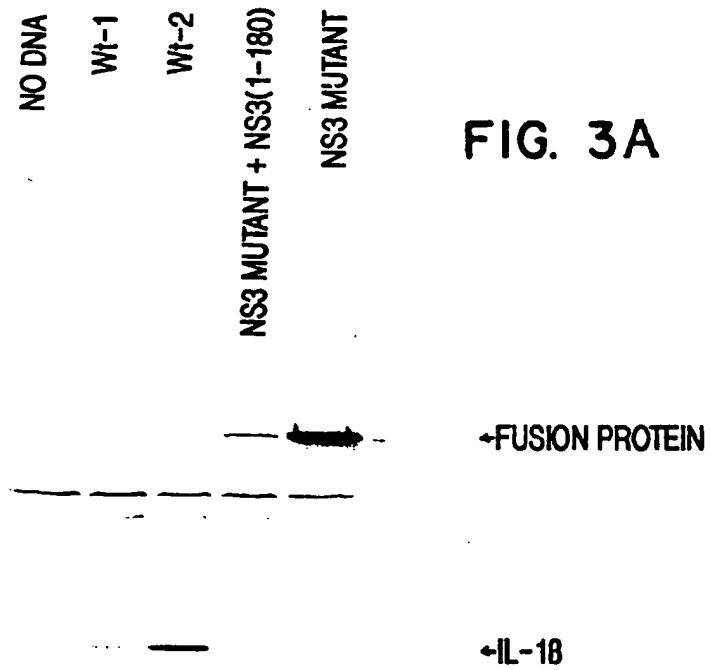


FIG. 3B

4 / 7

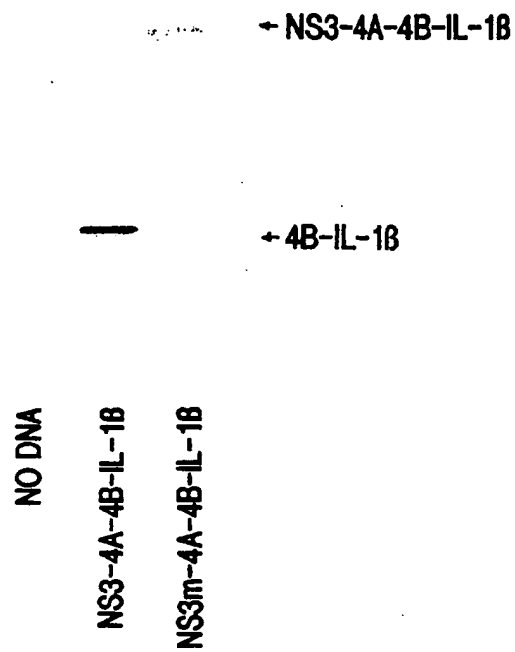


FIG. 4

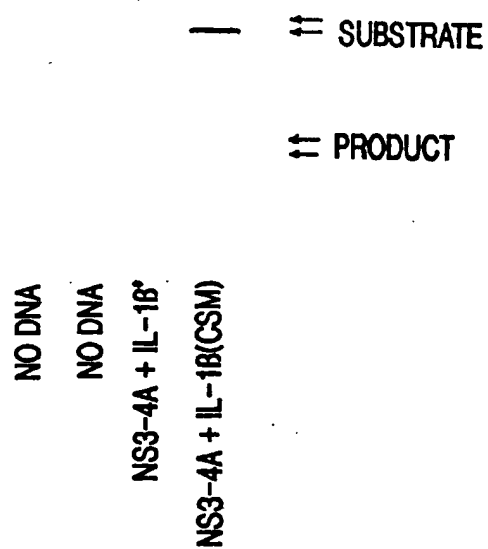


FIG. 5

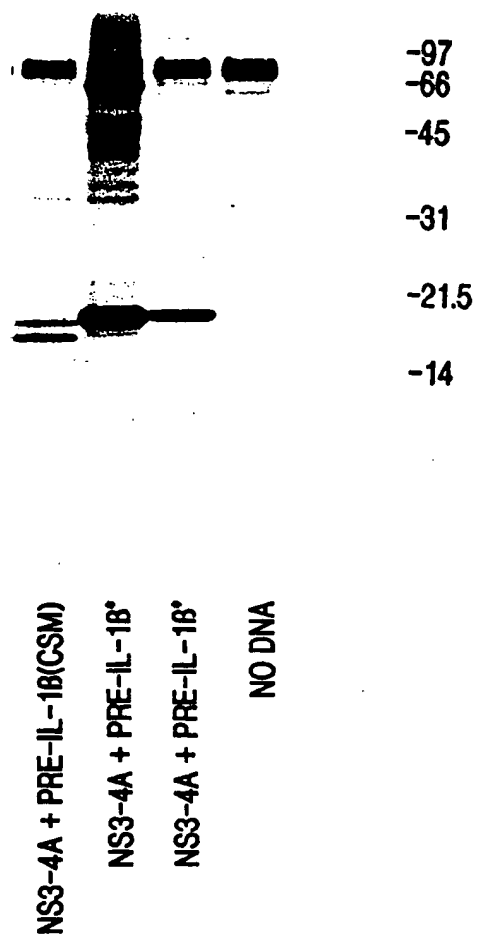


FIG. 6

7 / 7

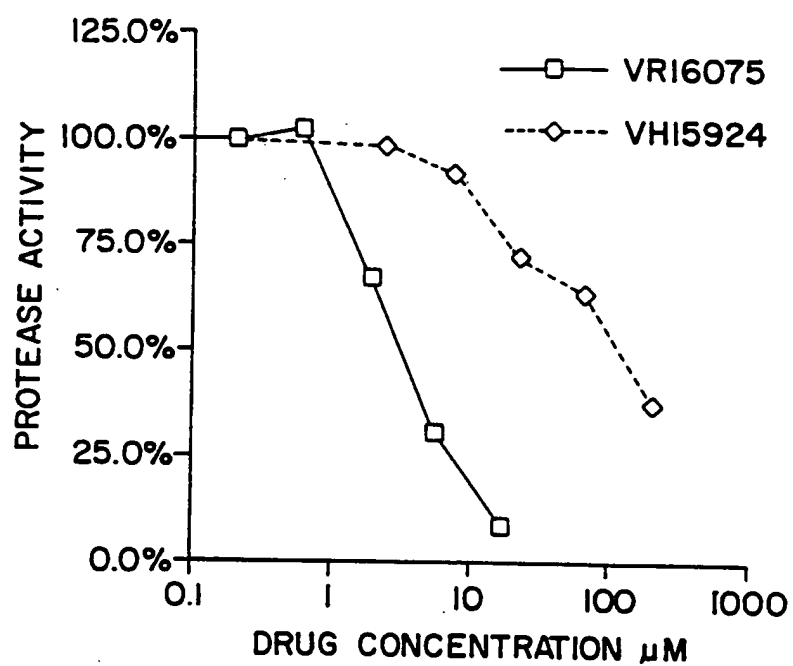


FIG. 7

INTERNATIONAL SEARCH REPORT

International Application No.
PC./US 96/06070

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12Q1/02 C12Q1/37 C12N5/10 C12N15/25

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12Q C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|---|----------------------------------|
| X | WO,A,95 02065 (UNIV COLORADO) 19 January 1995 see page 6, line 15 - page 8, line 2 see page 9, line 12 - line 15 see page 9, line 16 - line 21 see page 52, line 12 - line 17 | 1-3,6, 11,12, 18,20-22 |
| Y | see claims 29,30 --- | 4,5, 7-10, 13-17, 19,23 |
| X | WO,A,93 01305 (BALINT ROBERT) 21 January 1993 see the whole document --- | 1-3,11, 12,20 |
| | --- | |
| | -/- | |

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

8 July 1996

Date of mailing of the international search report

20.08.96

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax (+31-70) 340-3016

Authorized officer

Hoekstra, S

INTERNATIONAL SEARCH REPORT

International Application No
PC./US 96/06070

| C(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT | | |
|---|---|----------------------------------|
| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
| X | PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 88, no. 14, 15 July 1991, pages 5979-5983, XP000248161 LIEBIG H D ET AL: "PROTEINASE TRAPPING: SCREENING FOR VIRAL PROTEINASE MUTANTS BY A COMPLEMENTATION" see the whole document | 4,11,12, 16,18,20 |
| Y | see claims | 5,17 |
| X | --- GENE (1991), 97(2), 253-8 CODEN: GENED6;ISSN: 0378-1119, 1991, XP002007744 PECCEU, F. ET AL: "Human interleukin 1.beta. fused to the human growth hormone signal peptide is N-glycosylated and secreted by Chinese hamster ovary cells" see the whole document | 11, 13-15,20 |
| X | --- WO,A,90 10075 (NOVONORDISK AS) 7 September 1990 see claim 34 see page 1 - page 4 | 11 |
| X | --- WO,A,95 02059 (NOVONORDISK AS ;CHRISTIANSEN LARS (DK); PETERSEN JENS G LITSKE (DK) 19 January 1995 see the whole document | 11 |
| Y | --- DE,A,38 19 846 (WOLF HANS PROF DR DR) 14 December 1989 see the whole document see column 4, line 25 - line 45 | 4,5, 7-10, 13-17, 19,23 |
| A | --- EP,A,0 421 109 (AMERICAN CYANAMID CO) 10 April 1991 see the whole document | 1-10 |
| A | --- J. BIOL. CHEM. (1993), 268(29), 22170-4 CODEN: JBCHA3;ISSN: 0021-9258, 1993, XP002007745 SIDERS, WILLIAM M. ET AL: "Characterization of the structural requirements and cell type specificity of IL-1.alpha. and IL-1.beta. secretion" | 1-23 |
| A | --- WO,A,91 15575 (CHIRON CORP) 17 October 1991 see the whole document | 5,17 |
| | ----- | |

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PL, US 96/06070

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---------------------|--|--|
| WO-A-9502065 | 19-01-95 | US-A- 5413914 AU-B- 7254694 | 09-05-95 06-02-95 |
| WO-A-9301305 | 21-01-93 | AU-B- 2377392 | 11-02-93 |
| WO-A-9010075 | 07-09-90 | AT-T- 110414 AU-B- 624694 AU-B- 5261290 CA-A- 2050336 DE-D- 69011853 DE-T- 69011853 EP-A- 0461165 ES-T- 2062514 JP-T- 4504846 PL-B- 163532 US-A- 5395922 US-A- 5510249 US-A- 5514585 | 15-09-94 18-06-92 26-09-90 04-09-90 29-09-94 15-12-94 18-12-91 16-12-94 27-08-92 29-04-94 07-03-95 23-04-96 07-05-96 |
| WO-A-9502059 | 19-01-95 | AU-B- 7122194 PL-A- 312436 ZA-A- 9404912 | 06-02-95 29-04-96 27-03-95 |
| DE-A-3819846 | 14-12-89 | NONE | |
| EP-A-0421109 | 10-04-91 | AU-B- 3392893 AU-B- 636383 AU-B- 6208490 CA-A- 2024277 JP-A- 3164196 | 20-05-93 29-04-93 14-03-91 12-03-91 16-07-91 |
| WO-A-9115575 | 17-10-91 | AU-B- 7675491 EP-A- 0527788 US-A- 5371017 | 30-10-91 24-02-93 06-12-94 |